PATENT ABSTRACTS OF JAPAN

(11)Publication number :   06-348696

(43)Date of publication of application : 22.12.1994

(51)Int.Cl.	G06F 15/20

|		
(21)Application number : 06-113093	(71)Applicant :	XEROX CORP
(22)Date of filing : 26.05.1994	(72)Inventor :	PEDERSEN JAN O
		TUKEY JOHN W

(30)Priority

Priority number : 93 71114   Priority date : 03.06.1993   Priority country : US

(54) AUTOMATIC IDENTIFICATION METHOD

(57)Abstract:

PURPOSE: To provide an automatic identification method which ranks the document sentences to produce the abstracts or to edit the documents through detection of the words and/or phrases that show the emphases.

CONSTITUTION: A document is inputted to a computer (200), and the first sentence is extracted (202). The sentence is scored by a scoring method (204). The computer checks and confirms whether all sentences are scored (206). If an unscored sentence is confirmed, a flow returned to the step 202 to extract the next sentence. If all sentences are scored, the flow proceeds to a step 208 where a prescribed number of sentences are selected based on the scores of every sentence. The selected documents are constructed (arrayed) in a sequence shown by the sentences, i.e., in a sequence that is not equal to the score sequence (210). Otherwise, an operator reconstructs the sentences or the sentences are constructed based on their scores. These sentences are displayed on a terminal or outputted by a printer, etc., (212).
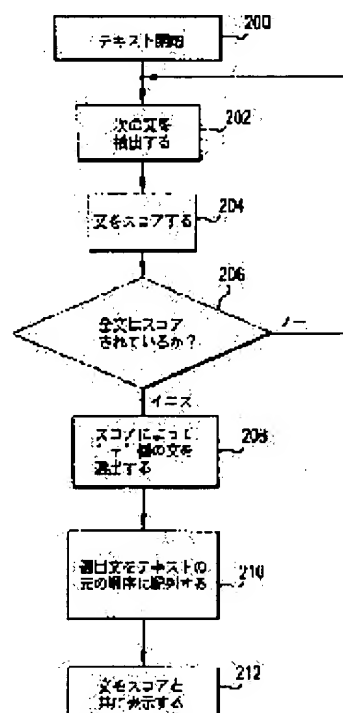
LEGAL STATUS

[Date of request for examination]   11.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]
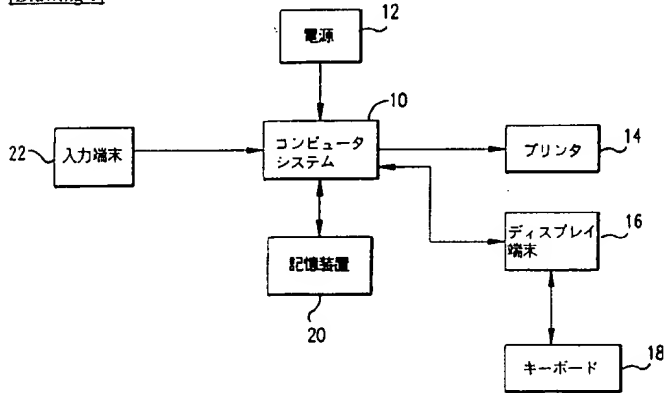
[Date of extinction of right]

* NOTICES *

DRAWINGS

[Drawing 1]

[Drawing 7]

" a "	"all"	"an"	"any"	"any"
"each"	"each"	"her"	"his"	"its"
"ordinary"	"other"	"our"	"particular"	"some"
"similar"	"some"	"some"	"sort"	"special"
"specially"	"the"	"their"	"these"	"those"
"totally"	"usually"	"which"	"whose"	

[Drawing 2]
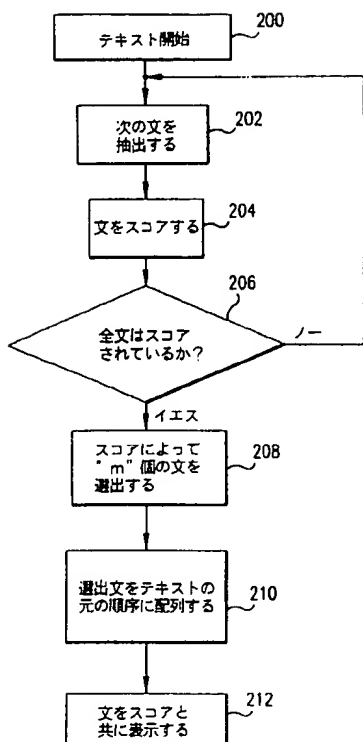
テキスト開始 — 200

次の文を抽出する — 202

文をスコアする — 204

全文はスコアされているか? — 206
ノー
イエス

スコアによって"m"個の文を選出する — 208

選出文をテキストの元の順序に配列する — 210

文をスコアと共に表示する — 212

[Drawing 3]

抽出された文中の各語をチェックし、全ストップ・ワードにフラグを立てる。 — 300

ランの数(m)を数える — 310

各ストップ・ワードをチェックし、全ヴァニッシュ・ワードにフラグを立てる — 302

ヴァニッシュ・ワードでないストップ・ワードの数(n)を数える — 304

3を加える[n+3] — 306

ストップ・ワードの数をランの数で割ることによってスコアする (n+3)／m=スコア — 308
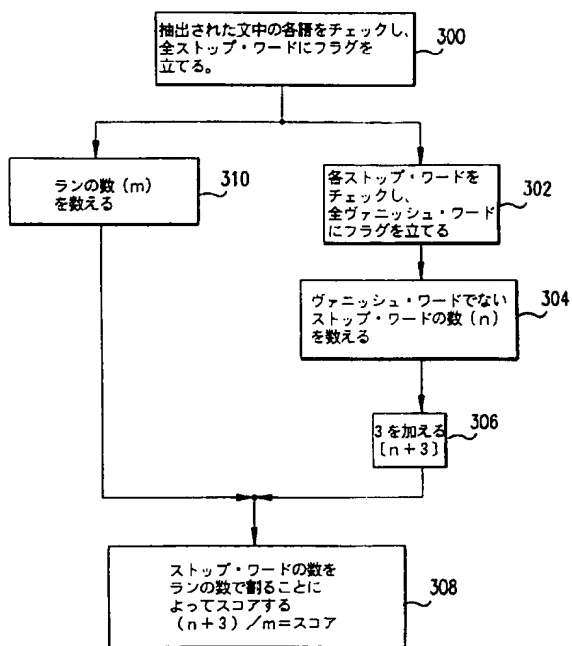
[Drawing 4]

"l"	"lnc"	"lld"	"o"	"able"
"about"	"above"	"across"	"act"	"actual"
"actually"	"after"	"afterword"	"again"	"against"
"ago"	"ahead"	"all"	"allow"	"almost"
"alone"	"along"	"already"	"also"	"although"
"always"	"am"	"among"	"amongst"	"an"
"and"	"another"	"any"	"anyhow"	"anyone"
"anything"	"anyway"	"anywhere"	"approach"	"appropriate"
"are"	"argue"	"arise"	"around"	"as"
"aside"	"assume"	"at"	"avail"	"avoid"
"away"	"awhile"	"b"	"back"	"bad"
"be"	"because"	"become"	"before"	"beforehand"
"begin"	"behave"	"behind"	"being"	"below"
"beside"	"best"	"better"	"between"	"beyond"
"big"	"both"	"bring"	"but"	"by"
"c"	"can"	"care"	"cause"	"certain"
"certainly"	"clear"	"clever"	"co"	"come"
"comment"	"completely"	"consider"	"considerate"	"convenient"
"could"	"course"	"cove"	"d"	"de"
"decide"	"decision"	"despite"	"didn"	"different"
"difficult"	"down"	"during"	"e"	"each"
"earlier"	"effect"	"eg"	"either"	"else"
"elsewhere"	"end"	"enough"	"entirely"	"especially"
"essential"	"etc"	"even"	"eventual"	"ever"
"every"	"everybody"	"everyone"	"everything"	"everywhere"
"evident"	"exact"	"example"	"except"	"exclusively"
"exist"	"expect"	"explain"	"extra"	"extract"
"extreme"	"f"	"fact"	"far"	"few"
"find"	"first"	"follow"	"following"	"for"
"fortunate"	"forward"	"found"	"from"	"full"
"fully"	"further"	"furthermore"	"g"	"gain"
"gether"	"gave"	"generally"	"get"	"give"
"given"	"go"	"going"	"gone"	"good"
"get"	"great"	"guess"	"h"	"ha"

[Drawing 5]

"hereafter"	"hereby"	"herein"	"hereupon"	"hers"
"herself"	"him"	"himself"	"his"	"hold"
"hope"	"how"	"however"	"i"	"idea"
"ideal"	"ie"	"if"	"ified"	"imagine"
"immediate"	"important"	"in"	"inc"	"include"
"indeed"	"instead"	"into"	"it"	"its"
"itself"	"j"	"just"	"k"	"know"
"l"	"la"	"large"	"last"	"later"
"latter"	"lay"	"least"	"leave"	"less"
"lesser"	"lest"	"let"	"lie"	"like"
"likely"	"little"	"ll"	"look"	"lot"
"low"	"lld"	"m"	"many"	"mark"
"may"	"maybe"	"me"	"mean"	"meanwhile"
"mere"	"might"	"mild"	"more"	"moreover"
"most"	"mostly"	"much"	"must"	"my"
"myself"	"n"	"none"	"near"	"nearly"
"necessary"	"nee"	"neither"	"nevertheless"	"next"
"nice"	"no"	"non"	"none"	"noone"
"nor"	"notably"	"note"	"notion"	"notwithstanding"
"now"	"nowhere"	"o"	"obvious"	"occur"
"of"	"off"	"often"	"ok"	"okay"
"on"	"once"	"one"	"only"	"onto"
"or"	"other"	"otherwise"	"ought"	"our"
"ourselves"	"out"	"over"	"own"	"p"
"particular"	"per"	"perfect"	"perhaps"	"please"
"plenty"	"possible"	"preferably"	"pretty"	"previous"
"probably"	"problem"	"produce"	"prompt"	"proper"
"propos"	"propose"	"put"	"q"	"quick"
"quite"	"r"	"rather"	"re"	"ready"
"really"	"reasonable"	"reasonably"	"regard"	"relatively"
"relevant"	"respect"	"s"	"some"	"saw"
"say"	"see"	"seem"	"send"	"sense"
"serious"	"serve"	"several"	"shall"	"she"
"shortly"	"should"	"show"	"similar"	"simple"
"simply"	"since"	"slight"	"slightly"	"slow"
"so"	"some"	"somebody"	"somehow"	"someone"
"something"	"sometime"	"sometimes"	"somewhat"	"somewhere"
"soon"	"sorry"	"specific"	"specifically"	"stay"

[Drawing 6]

"thereby"	"therfore"	"therein"	"thereupon"	"these"
"they"	"thing"	"this"	"those"	"though"
"thought"	"through"	"throughout"	"thru"	"thus"
"til"	"till"	"time"	"tire"	"to"
"together"	"too"	"total"	"toward"	"treat"
"tried"	"trouble"	"truly"	"try"	"two"
"u"	"under"	"unless"	"unlikely"	"until"
"up"	"upon"	"us"	"use"	"useful"
"useless"	"usual"	"usually"	"v"	"various"
"ve"	"very"	"via"	"view"	"w"
"way"	"we"	"well"	"what"	"whatever"
"when"	"whence"	"whenever"	"where"	"whereafter"
"whereas"	"whereby"	"wherein"	"whereupon"	"wherever"
"whether"	"which"	"while"	"whither"	"who"
"whoever"	"whole"	"whom"	"whose"	"whosever"
"why"	"will"	"with"	"within"	"without"
"would"	"x"	"y"	"yes"	"yet"
"you"	"your"	"yours"	"yourself"	"yourselves"
"z"				

[Drawing 8]
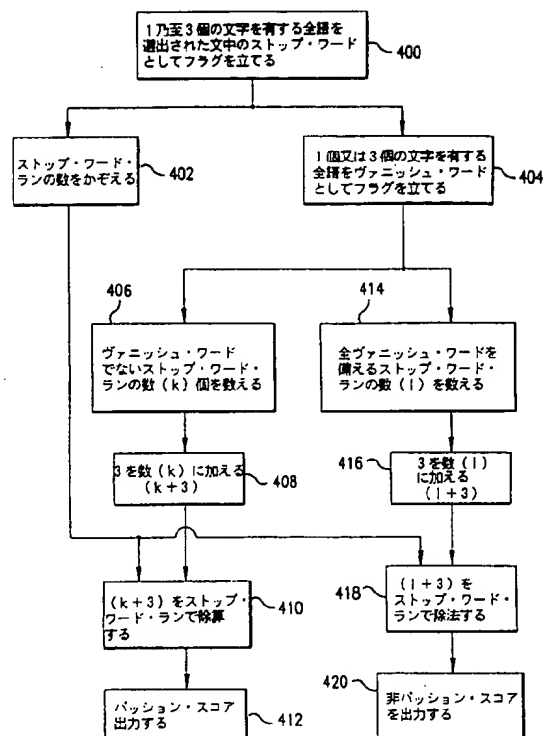
400 1乃至3個の文字を有する全語を 選出された文中のストップ・ワード としてフラグを立てる

402 ストップ・ワード・ ランの数をかぞえる

404 1個又は3個の文字を有する 全語をヴァニッシュ・ワード としてフラグを立てる

406 ヴァニッシュ・ワード でないストップ・ワード・ ランの数（k）個を数える

414 全ヴァニッシュ・ワードを 備えるストップ・ワード・ ランの数（l）を数える

408 3を数（k）に加える （k＋3）

416 3を数（l） に加える （l＋3）

410 （k＋3）をストップ・ ワード・ランで除算 する

418 （l＋3）を ストップ・ワード・ ランで除法する

412 パッション・スコア 出力する

420 非パッション・スコア を出力する

[Drawing 9]

500 テキストの始まり

502 次の文を抽出する

504 全部のストップ・ワード・ ランにフラグを立てる

506 未処理のランが 存在するか？
ノー
イエス

508 次のラン を調べる

510 このランは ロングか？
ノー
イエス

512 ショート・ラン カウンタを1だ け増分する

514 ショート・ ラン・カウンタは 0であるか？
ノー
イエス

516 エンド・レジスタを 現在のランのすぐ前の 語の位置に設定する

520 スタート・レジスタを 現在のランの直ぐ次に 続く語の位置に設定する

518 ショート・ラン・カウンタ の内容及びレジスタの位置 を記憶する

522 ショート・ラン・ カウンタを0に設定する

* NOTICES *

```
Japan Patent Office is not responsible for any
damages caused by the use of this translation.
```
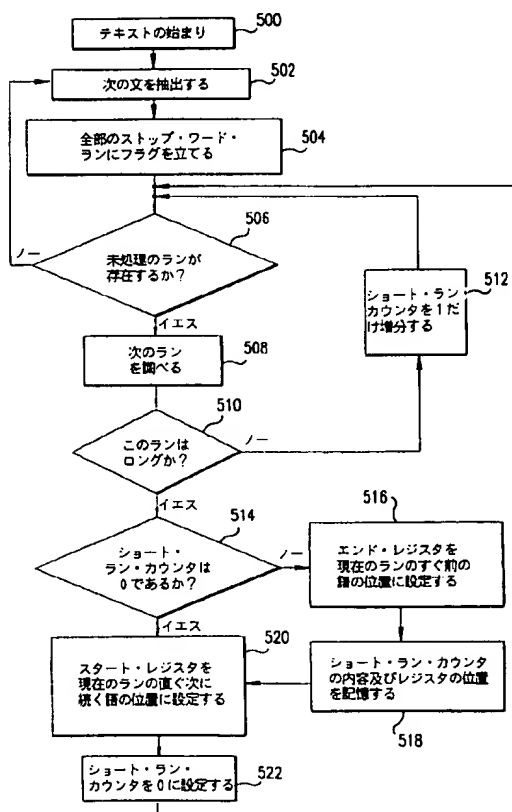
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the schematic drawing showing the standalone version computer by which this invention may be used.
[Drawing 2] It is a flow Fig. showing one method of summarizing a document automatically.
[Drawing 3] It is the flow Fig. showing the approach for carrying out scoring of the sentence (or other blocks of a text) according to the 1st example.
[Drawing 4] It is the sample of a stop list.
[Drawing 5] It is the sample of a stop list.
[Drawing 6] It is the sample of a stop list.
[Drawing 7] It is the sample of a VANISSHU list.
[Drawing 8] It is the flow Fig. showing the approach for carrying out scoring of the sentence (or other blocks of a text) according to the 2nd example.
[Drawing 9] It is the flow Fig. showing the approach for carrying out scoring of the sentence (or other blocks of a text) according to the 3rd example.

[Translation done.]

h   g   cg   b   eb cg   e   e

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] This invention relates to the approach and equipment for creating the abstract (abstract) of a document (document). Furthermore, through detection of the word (WORD) and/or phrase (phrase) which show emphasis, this invention can be used for a detail in order to create an opium poppy with a rank, and an abstract for the sentence (sentence) of a document or to edit a document.
[0002]
[Description of the Prior Art] Since it can judge about the appropriateness of a document, without reading the whole document, a reader can save time amount by the abstract of a document. There are two types of abstracts. The abstract of the 1st type summarizes the main contents of the document. Although the abstract of the 2nd type does not summarize a document, it explains the general theme of a document instead.
[0003] Generally a document abstract is needed for a formal publication. However, all the abstracts to which no (prepared first) documents restrict having the abstract but for which they are further prepared manually by people are also appropriate. Therefore, the practical automatic configuration of a useful document abstract is needed.
[0004] Itself of the automatic document abstract is clearly useful, and it can also become a bigger structure-of-a-system element. For example, a document (reference) retrieval system is a system which generally moves to all the words in a document based on a document, i.e., 2 or 3 words, of a word based on an inquiry. It is useful to decrease the step size of this jump by moving to a document abstract instead by inquiry. Especially, it contracts appropriately, and by summarizing, arbitrary long documents are compressed and can suit on 1 screen of a screen.
[0005] An automatic extracting system identifies the word which exists in an ordinary text (text), and the word which exists in a title or a header. An ordinary text can receive standard vocabulary weighting and the special word of a title or a header can receive special processing based on the vocabulary location (location) of a specific document. In some systems, the sentence of the beginning of every knot (paragraph) is only chosen. In an option, special processing is performed even to a word with high frequency, the word which is not used rarely, a specific phrase, or a specific knot. Next, the score (grading) of each sentence or knot is carried out by the frequency of a word or a phrase. Such an abstract creation technique is described by "automatic text-manipulation (Automatic Text Processing)" (the Gerald ape ton, Addison - WEZURI, 1989).
[0006]
[Means for Solving the Problem and its Function] Only a very simple language analysis is needed in this invention. One side is accomplished in small size and the comparison of the suitable word of the text word of an example and a suitable word accomplishes another side by two Ward Liszt of inside size.
[0007] According to this invention, a document is treated as a sequence of each sentence (or block of the text over two or more sentences (assembly)). While obtaining contraction of the text for which it asks, the subset of the sentence (or block) which serves as a useful representation sample of a document is

h   g   cg   b   eb cg   e   e

chosen. Since it is guaranteed that an output is well formed like an input, this strategy avoids a language generation problem appropriately. However, a demand of usefulness asks for generating a subset still more attractive than the subset generated by fundamental approaches, such as selecting the sentence of the beginning of each knot, in order that these subsets may generate selecting a document as arbitration, or a subset.

[0008] This invention chooses the sentence pn individual as which the document which consists of n sentences (constituted as a word in the sentence which is generally in a knot) was searched, and the number of requests was elected. In this case, p shows a contraction (reduction) factor. stop (STOP) Liszt compensated by two words Liszt and VANISSHU (empty is sufficient) (VANISH) Liszt -- desirable -- ** et al. -- ** -- it is used for a search. When stop Liszt has a word, the word is stop Ward, and when stop Liszt does not have a word conversely, the word is a contents (content) word. Since a certain kind of stop word exists also in VANISSHU Liszt, it is also VANISSHU (it disappears) Ward. A stop and VANISSHU Liszt are extensible by arranging a word to a suffix or the equivalence class (namely, word stem) based on analysis of morphology. Even in a document or each knot, and a sentence, the frequency of each word is recordable in order to give next election. After setting a flag to stop Ward, a sentence forms the mutual run (ream) of a contest word with stop Ward. The number of stop Ward Laon is used in order to carry out the score of the sentence. Based on a score, since an abstract is constituted, a certain sentence can be used.

[0009] the procedure indicated here is useful to a matter important for a writer, or the detection of a sentence which is related to passion (many ** -- sensitive) still more generally.

[0010] In order to make it a practical use target, the strategy which is completely dependent on natural language understanding should not be used preferably. It depends on the skewness of the direction for use of stop Ward or a short word for the technique used for this invention. This clear direction for use of stop Ward acts as what is replaced with the structure in a text (text) where the back is deeper. the sentence identified by the technique of this invention as a very passionate (many ** -- sensitive) sentence is a sentence which describes the matter in which a writer has strong feeling (feeling), such as either [joy,] praise or sadness.

[0011] The step which the mode of this invention is an approach for discriminating the field of a text from the electronic filing document which uses a digital computer automatically, and identifies stop Ward in a document, The step which determines whether it is VANISSHU Ward any of stop Ward which were identified they are, It has the step which answers stop Ward and VANISSHU Ward in each field, and carries out scoring of the field of a document, and the step which identifies a number of fields with which the document was beforehand defined based on the score of a field.

[0012]

[Example] Reference of drawing 1 shows the schematic diagram showing standalone version computer system. A power source 12 supplies power to the computer system 10 which has CPU (not shown) and memory (not shown). The input terminal 22 loads a document text to a computer 10. When some examples of an input terminal are raised, there are a manuscript scanner which is equipped with optical-character-recognition equipment (OCR), or is not equipped with optical-character-recognition equipment, a word processor, a floppy disk drive, a modem, etc. Storage 20 memorizes the score (grading) of a sentence with all the elected texts. The sentence by which the score was carried out can be outputted to a printer 14 and the display (display) terminal 16. Or an operator inputs a command for a keyboard 18, he can use it for changing a sentence into an abstract etc. (if required).

[0013] A document is divided into two or more fields of the text (text) by which a score is carried out in the following suitable examples. Each field can become the fragmentation (fragment; in part) of a sentence and a sentence, the block (assembly) of a sentence (the range was limited in order to obtain the minimum total which asks for stop Ward's run (ream) (run) probably), a half-paragraph, or a paragraph. Stop Ward is identified by either comparing a text with stop Liszt, or setting a flag into the word of a certain length (die length) as stop Ward. VANISSHU Ward who is stop Ward's subset is identified by either comparing stop Ward with VANISSHU Liszt, or setting a flag to stop Ward of the length made into VANISSHU Ward. Next, depending on the number of stop Ward, VANISSHU Ward, and the stop

h   g cg b   eb cg e e

Ward groups in each field (group), the score of the field (for example, sentence) of two or more documents is carried out. A number with a document of fields are elected based on the score of each field. A user can create an abstract based on these elected fields, or an abstract is created automatically and he gets.

[0014] Reference of <u>drawing 2</u> shows the flow Fig. showing how to summarize a document automatically. First, a document is inputted into a computer (step 200). The first sentence (or block of a word) is extracted (step 202). The score (grading) of the sentence is carried out by the scoring (further explained to detail below) approach (step 204). In order to check whether it is the no to which the score of the whole sentence was extracted and carried out, a computer checks (step 206). (inspection) Furthermore, if a text exists, a flow will return to step 202, in order to extract the following sentence. Moreover, if the score of all the texts was carried out, it will progress to step 208 and the sentence of a predetermined number will be selected based on the score for every sentence. An election document is constituted by the sequence shown in the text, i.e., the sequence which is not the sequence of a score, (step 210). (array) Or an operator reconfigurates a sentence, or a sentence is constituted according to a score and gets. These sentences are displayed on a terminal 16 (score which corresponds although it is arbitrary), or when that is not right, they are outputted by printer 14 grade (step 212).

[0015] the 1st suitable example -- a stop/VANISSHU (STOP/VANISH) -- it can be shown as law. This approach detects emphasis by measuring stop Ward's use pattern. this example is based on the observation on experience in which it is in the inclination for a passionate (many ** -- sensitive) sentence not to be equipped with stop Ward's long run. therefore, the score which measures the average stop Ward Laon length of a given sentence (or block of a word) -- reverse ** indicator (inverted indicator; reverse index) of passion (many ** -- sensitive) ****** -- it acts. A sentence is classified into the ascending order of this score (sort), it is the minimum scoring sentence of a pxn individual, and the sentence preferably arranged in order of the original reading is chosen as an abstract.

[0016] Especially, it is the i-th sentence Si. Suppose that it can express with the sequence of the stop WORD run Si and j. The following formulas are good for defining the index about passion.

[0017]

[Equation 1]

$$score(S_i) = \frac{\Sigma_j |S_{i,j}| + K}{|S_i|}$$

$$= Ave(|S_{i,j}|) + K/|S_i|$$

[0018] In this case, $|Si|$ is the number of stop Ward Laon in a sentence, $|\{Si, j\}|$ is the number of stop Ward j-th in stop Ward Laon, and a little term K/|Si| imposes a penalty (fine) on the sentence which has stop Ward Laon which is not (average run length is originally more changeable). K is a fraction (fraction). A general value is set to 3 although it can grow into all the included numbers.

[0019] CP Suppose that it is the minimum scoring sentence of pn individual. Next, an abstract is Cp of index sequence. It becomes. namely, Cp = {Si1, Si2, ..., Sipn} -- this case -- i1 <i2 -- < ... it is <ipn.

[0020] Discernment of an important sentence is improvable by offering short (short) Liszt of VANISSHU Ward who does not contribute to stop Ward Laon length. That is, |Si and j | are stop Ward Laon Si and j which is not on VANISSHU Liszt. It is changed so that the number of VANISSHU Ward may be subtracted from the total of a word (count), i.e., the number of stop Ward. Count (count) can be set to 0. VANISSHU Liszt is prepared so that the word which only turns to a related word (a focus is carried out), or individualizes it into it (PASONA rise) may not be counted (to arbitration). For example, including all the closing set words (a determiner, a pronoun, preposition, etc.), stop Liszt contains "a", "an", "its, and "their", and VANISSHU Liszt is "of further. a", "of an", "of its" and "oftheir" will be counted, respectively as stop Ward Laon of die length (it is equivalent to of) 1. Stop Ward Laon which includes VANISSHU Liszt's word completely is counted as stop Ward Laon of die length 0.

[0021] Another possible example performs an above-mentioned approach in the aperture (Ward window) (or block of a word) of the word to which each contains either the constant of a word, or the

h   g   cg   b   eb   cg   e   e

constant of stop Ward Laon in a text, and includes electing the block which has the minimum stop Ward Laon score. These blocks are crossed to one or more sentences it is considered that are the parts (segment) of the text in which the whole was emphasized. Or the sentence exceeding the specific fraction to which length belongs to such a block is elected, and it gets.

[0022] Reference of drawing 3 shows the flow Fig. for carrying out scoring (step 204 of drawing 2) of the sentence according to the 1st suitable example. As mentioned above, this processing can also be used for the change based on a sentence based on the block of a word. Each word in the extracted sentence is compared with stop (STOP) Liszt's word (step 300). Drawing 4 thru/or stop Liszt typical to 6 are shown. As for stop Ward, a flag is set. Each stop Ward is compared with VANISSHU (VANISH) Liszt's Ward (step 302). Each VANISSHU Ward can build a flag (additionally). VANISSHU Liszt is the arbitrary descriptions of suitable this example. Since VANISSHU Ward is a common language which carries out the PASONA rise (individualization) of related Ward, he is not counted as what contributes to the length of stop Ward Laon. Typical VANISSHU Liszt is shown in drawing 7 . Stop Ward who was not able to stand a flag as VANISSHU Ward is counted, and it memorizes as n pieces (step 304). Number"3" is added and memorized by n (step 306). The addition of a number is arbitrary and the (it is not necessary to be an integer) numeric value can be changed. Increase of the minimum size of a sentence (or block) can decrease the need of adding the number part of a score.

[0023] The number of runs is counted (step 310). Stop Ward Laon is a block of adjoining stop Ward. Several m of a run is memorized. The number of stop Ward who is not VANISSHU Ward is divided by the number of stop Ward Laon in a sentence (step 308). the score with which it is obtained as a result (stop run length of the changed average) is a score of a sentence, is boiled for election of the most passionate sentence and used later.

[0024] An example of the scoring of the sentence which uses the approach of the 1st example for below is shown. The sentence of the following texts is : [0025] by which a score is carried out.

[External Character 1]

h   g   cg b   eb cg e e

The most important invention that will come out of the corporate research lab in the future will be the corporation itself. As companies try to keep pace with rapid changes in technology and cope with increasingly unstable business environments, the research department has to do more than simply innovate new products. It must design the new technological and organizational "architectures" that make possible a continuously innovating company. Put another way, corporate research must reinvent innovation.

At the Xerox Palo Alto Research Center (PARC) we've learned this lesson, at times, the hard way. Xerox created PARC in 1970 to pursue advanced research in computer science, electronics, and materials science. Over the next decade, PARC researchers were responsible for some of the basic innovations of the personalcomputer revolution-only to see other companies commercialize these innovations more

quickly than Xerox. (See the insert "PARC": Seedbed of the Computer Revolution.") In the process, Xerox gained a reputation for "fumbling the future" and PARC for doing brilliant research but in isolation from the company's business.

That view is one-sided because it ignores the way that PARC innovations have paid off over the past 20 years. Still, it raises fundamental questions that many companies besides Xerox have been struggling with in recent years: What is the role of corporate research in a business environment characterized by tougher competition and nonstop technological change? And how can large companies better assimilate the latest innovations and quickly incorporate them in new products?"

[0026] The score of each sentence of this text is extracted and carried out. The emphasis detection score corresponding to each sentence is indicated next to each sentence (Liszt). Scoring used the emphasis detection type using VANISSHU Liszt of arbitration. In the following sentences, as for each stop Ward Laon, an underline is drawn, and each VANISSHU Ward is shown by italic type.
[0027]
[External Character 2]

h   g   cg b   eb cg e e

2.8- *The* most important invention that will come out of *the* research lab in *the* future will be *the* corporation itself.

スコアは次のように計算される:
$$\frac{(15ストップ・ワード-4ヴァニッシュ・ワード)+3}{5ストップ・ラン}=2.8$$

1.7- As companies try to keep pace with rapid changes in technology and cope with increasingly unstable business environments, *the* research department has to do more than simply innovate new products.

1.6- It must design *the* new technology and organizational "architectures" that make possible a continuously innovating company.

3.5- Put another way, corporate research must reinvent innovation.

1.8- At *the* Xerox Palo Alto Research Center (PARC) we've learned this lesson, at times, *the* hard way.

1.8- Xerox created PARC in 1970 to pursue advanced research in computer science, electronics, and materials science.

1.8- Over *the* next decade, PARC researchers were responsible for some of *the* basic innovations of *the* personal-computer revolution-only to see other companies commercialize *these* innovations more quickly than Xerox.

2.5- (See *the* insert "PARC: Seedbed of *the* Computer Revolution.")

1.4- In *the* process, Xerox gained a reputation for "fumbling *the* future" and PARC for doing brilliant research but in isolation from *the* company's business.

2.6- That view is one-sided because it ignores *the way that* PARC innovations have paid off over *the* past 20 years.

1.9- Still, it raises fundamental questions that many companies besides Xerox have been struggling with in recent years: What is *the* role of corporate research in a business environment characterized by tougher competition and nonstop technological change?

2.2- And how can large companies better assimilate *the* latest innovations and quickly incorporate them in new products?

[0028] The minimum scoring sentence of this example 'In the process and Xerox gained a reputation It is for "fumbling the future" andPARC for doing brilliant research but in isolation from the company's business.'. Next, 'It must design the new technological and organizational "architectures"that make possible a continuously innovating company.' -- and 'As companies try to keep pace with rapid changes in technology and copewith increasingly unstable business environments and the research department has to do more than simply innovate new products. continue by the near score. Although these sentences may not tell all the themes of the fragment of this text, they are clear. [of it being one of the inside currently stated most passionately] An abstract is made from these elected sentences.

[0029] It is considered that the highest scoring sentence is un-passionate (it is not passionate). Since possibly the un-passionate sentence was inserted so that record (record) might be made perfect so that a

h   g   cg b   eb cg e e

text might be reinforced or, it offers a continuity or information. These sentences are useful to creation of an abstract.

[0030] In the 2nd suitable example, a passionate sentence is detected by investigating the Ward length (the die length of a word) to a change of discernment of a word. Therefore, although stop Liszt and VANISSHU Liszt are unnecessary, in addition, a word is classified as stop Ward or VANISSHU Ward. The compaction proposal (compact method) shown as the short (SHORT) approach defines Ward as follows.

[0031] - Stop Ward is alphabetic characters (letter) fewer than three pieces or it. It has. And - The whole of VANISSHU Ward is the word of one piece and three alphabetic characters.

The scoring of a sentence is : [0032] shown as follows.

[Equation 2]

$$\text{パッション・スコア(の不足)} = \frac{3 + \Sigma\,\delta(r_i)}{|\{r_i\}|}$$

$$\text{非パッション・スコア(の不足)} = \frac{3 + \Sigma\,1\text{-}\delta(r_i)}{|\{r_i\}|}$$

この場合、|{r_i}|は、ランの数を示す ; そして

$$\delta(r_i) = \begin{cases} 0, & \text{ラン i 中の全語がヴァニッシュ・ワードである場合} \\ i, & \text{その他の場合} \end{cases}$$

[0033] The most passionate sentence is a sentence (it runs short) of the lowest passion score. The un-most passionate sentence is a sentence (it runs short) of the lowest non-passion score.

[0034] Electing the un-most passionate sentence can fulfill the useful purpose by proposing that it edits, although it may not be useful as a path (means) to an epitome. Therefore, discernment of an un-passionate sentence can be used as an edit tool.

[0035] Reference of drawing 8 shows the flow Fig. which carries out scoring of the sentence according to the 2nd suitable example (step 204 of drawing 2). All words equipped with three or less alphabetic characters can set a flag as stop Ward (step 400). The number of stop Ward Laon is counted (step 402). All stop Ward that has one piece or three alphabetic characters can set a flag as VANISSHU Ward (step 404). The number of stop Ward Laon equipped with the word which is not VANISSHU Ward is counted, and it memorizes as K pieces (step 406). K grand totals of +3 are memorized (step 408). it is arbitrary like a previous example to add 3 (or -- all -- others -- a number). Furthermore, if a number is removed completely, the grand total of two scores (passion and non-passion) will be set to 1. In this specific example, only one score (passion or non-passion) is needed.

[0036] It is determined by breaking the result from which the passion score was obtained at step 408 by the result obtained at step 402 (step 410). The output of step 412 is the passion score of the elected sentence.

[0037] A non-passion score is determined as follows. Each stop Ward Laon equipped with all VANISSHU Ward is counted, and it memorizes as l (step 414). Several 3 is added to the total of all stop Ward having all VANISSHU Ward (step 416). The result of the addition is memorized. It is determined by breaking the result from which the non-passion score was obtained at step 416 by the result obtained at step 402 (step 418). The output of step 420 is a non-passion score about the sentence elected.

[0038]

[External Character 3]

h   g cg b   eb cg e e

この方法は、２つの異なるテキストを有する以下の２つの文書に対して実行される：

1) <u>Research That Reinvents the Corporation</u>, Brown, John Seely, Harvard Business Review, Jan/Feb 1991, pp102-111. (The text has 236 sentences in 55 paragraphs or headings).

１）　企業復興研究

（Brown, John Seeley 、Harvard Business Review, Jan/Feb 1991, pp102-111.）
（テキストは、５５節又は項目中に２３６文を有する）。

2) <u>Transcript of the Remarks in Moscow</u>, New York Times, 21 December 1990, p. A7. (Edward Shevardnadze's resignation speech which is translated by the BBC) (The speech has 65 sentences in 30 paragraphs).

２）　モスクワの批評のトランスクリプト（写し）

（New York Times. 21 December 1990. p. A7.）
（ＢＢＣ放送によって翻訳されたEdward Shevardnadze's の辞任スピーチ）
（スピーチは３０節中に６５文を有する）。

[0039] The sentence from which the following was extracted shows the sentence of eight highest scoring by which the score was carried out as the passion in each text, and non-passion in order of the original text.
[0040] The difference between two sets selected out of each text is clear.
[0041] : to which the score of the following sentences was carried out as six most passionate sentences in Brown's editorial (as for stop Ward, an underline is drawn and VANISSHU Ward is shown by italic type.) [0042]
[External Character 4]

h　　　　　　　g　cg　b　　　eb cg　e　e

- Research must "coproduce" *new* technologies *and* work practices <u>by</u> developing with partners throughout *the* organization <u>a</u> shared understanding *of why* these innovations *are* important

パッション・スコアは次のように決定される

$$3 + ヴァニッシュ・ワードではない2ストップ・ワード・ラン = 0.71 \over 7ストップ・ワード・ラン$$

非パッション・スコアは次のように決定される

$$3 + 全ヴァニッシュ・ワードを備える5ストップ・ワード・ラン = 1.14 \over 7ストップ・ワード・ラン$$

- RIC is an expert system inside the copier that monitors the information technology controlling the machine and, using some artificial-intelligence techniques, predicts when the machine will next break down.
- Recently, Xerox introduce its most versatile office machine ever - a product that replaces traditional light-lens copying techniques with "digital copying," where documents are electronically scanned to create an image stored in a computer, then printed out whenever needed.
- They are storehouse of past problems and diagnoses, a template for constructing a theory about the current problem, and the basis for making an educated stab at a solution.
- The document was "unfinished" in the sense that the whole point of the exercise was to get the viewers to complete the video by suggesting their own ideas for how they might use the new technology and what these new uses might mean for the business.
- The Express team is exploring ways to use core technologies developed at PARC to help the pharmaceutical company manage the more than 300,000 "case report" forms it collects each year.

[0043] In Brown's editorial, the following sentence was ranked as eight un-most passionate sentences.
[0044]
[External Character 5]

h   g   cg   b   eb cg   e   e

- Still, it raises fundamental questions that many companies besides Xerox have been struggling with in recent years: What is the role of corporate research in a business environment characterized by tougher competition and nonstop technological change?

パッション・スコアは次のように計算される:

$$\frac{3+ヴァニッシュ・ワードではない6ストップ・ラン}{7ストップ・ワード・ラン}=1.29$$

非パッション・スコアは次のように計算される:

$$\frac{3+全ヴァニッシュ・ワードを備える1ストップ・ラン}{7ストップ・ワード・ラン}=0.57$$

- As RIC collects information on the performance of our copiers - in real-world business environments, year in and year out - we will eventually be able to use that information to guide how we design future generations of copiers.

- In effect, technology will become so flexible that users will be able to customize it evermore precisely to meet their particular needs - a process that might be termed "mass customization."

- People use procedures to understand the goals of a particular file has to contain in order for a bill to be paid - not to identify the steps to take in order to get from here to there.

- In most cases, ideas generated by employees in the course of their work are lost to the organization as a whole.

- We thought of the unfinished document as a "conceptual envisioning experiment" an attempt to imagine how a technology might be used before we started building it.

- We are also involved in initiatives to get managers far down in the organization to reflect on the obstacles blocking innovation in the Xerox culture.

- One step in this direction is an initiative of Xerox's Corporate Research Group (of which PARC is a part) known as the Express project.

[0045] These 14 sentences (six most passionate sentences and eight un-most passionate sentences) were selected out of 256 sentences. It can perform easily creating an abstract from these elected sentences the reader could understand the fundamental propositions to be behind the editorial at anyone.
[0046] In the speech of Shevardnadze, the following sentences were ranked as eight most passionate sentences.

h   g   cg   b   eb cg   e   e

[0047]
[External Character 6]
In the Shevardnadze speech the following sentences were

- Second, I have explained repeatedly, and Mikhail Sergeyevich spoke of this in his speech at the Supreme Soviet that the Soviet leadership does not have any plans - I do not know, maybe someone else has some plans, some group - but official bodies, the Ministry of Defense - charges are made that the Foreign Minister plans to land troops in the Persian Gulf, in the region.

- Is it an accident that two members of the Parliament make a statement saying that the Minister of Internal Affairs was removed successfully and that the time has come to settle accounts with the Foreign Minister?

- Because at the congress a real struggle developed, a most acute struggle, between the reformers and - I will not say conservatives, I respect the conservatives because they have their own views which are acceptable to society - but the reactionaries, precisely the reactionaries.

- And this battle, it must be stated bluntly, was won with merit by the progressive section, the progressive members, delegates, the progressively minded delegates to the congress.

- On comrade Lukyanov's initiative, literally just before the start of a meeting, a serious matter was included on the agenda about the treaties with the German Democratic Republic.

- Not one person could be found including the person in the chair to reply and say simple that this was dishonorable that this is not the way not how things are done in civilized states.

- I will not name the publications, all manner of publications that pamyat society - I add the pamyat society to these publications - but what statements: down with the Gorbachev clique.

- I nevertheless believe that the dictatorship will not succeed, that the future belongs to democracy and freedom.

[0048] In the speech of Shevardnadze, the following sentences were ranked as eight un-most passionate

h   g   cg   b   eb cg   e   e

sentences.
[0049]
[External Character 7]

- I have drawn up the text of such a speech, and I gave it to the secretariat, and the deputies can acquaint themselves with it - what has been done is the sphere of current policy by the country's leadership, by the President and by the ministry of Foreign Affairs, and how the current conditions are shaping up for the development of the country, for the implementation of the plans for our democratization and renewal of the country, for economic development and so on.

- In that case we would have had to strike through everything that has been done in recent years by all of us, by the whole country and by all of our people in the field of asserting the principles of the new political thinking.

- The third issue, I said there, and I confirm it and state it publicly, that if the interests of the Soviet people are encroached upon, if just one person suffers - wherever it may happen, in any country, not just in Iraq but in any other country - yes, the Soviet Government, the Soviet side will stand up for the interests of its citizens.

- And what is surprising, and I think we should think seriously: who is behind these comrades, and why is no one rebuffing them and saying that this is not so and that there are no such plans?

- Because many people think that the ministers who sit there or the members of the Government or the President, or someone else, are hired, and that they can do what they like with them.

- I would like to recall that it was against my will, without my being consulted, that my name, my candidacy was included for secret voting. And I had 800 against, 800 delegates voted against.

- No one knows what this dictatorship will be like, what kind of dictator will come to power and what order will be established.

- Let this be - and do not react and do not curse me - let this be my contribution, if you like, my protest against the onset of dictatorship.

[0050] This is Brown of said publication. Although the editorial was the text of a very different type, the abstract of this speech that has 65 sentences was summarized by 16 sentences. These eight elected passionate sentences express a lot of passion which a speaker pulls out in a speech. An un-passionate sentence adds the information about the contents and the background of a speech. The abstract which describes the fundamental idea of a speech in back is easily created from these 16 sentences.

[0051] This approach can be used also by using the Ward (not being discernment of word) shapes (class of the form of a word). For example, the output from Ward shape recognition equipment can be used in order to classify Ward in a sentence.

[0052] The 3rd example for detecting a passionate sentence looks for the long string (long train) of short stop Ward Laon. this approach -- long - short (LONG-SHORT) -- label attachment is carried out as law. Stop Liszt and VANISSHU Liszt of drawing 4 thru/or drawing 7 are preferably used for this example. Short stop Ward Laon is all Ward or the Ward group including one either equipped with one stop Ward or at least one VANISSHU Ward of stop Ward. Long stop Ward Laon is all other stop Ward Laon. A sentence (or block) including the long sequence of short stop Ward Laon is elected noting that it is a text for which it asks.

[0053] Reference of drawing 9 shows the flow Fig. for electing a sentence according to the 3rd suitable example (step 204 of drawing 2). A start register is arranged at the beginning of a text, and the Short Laon counter is set as 0 (step 500). The sentence of the beginning of a document is extracted (step 502). It is identified within the sentence from which all stop Ward Laon was extracted (step 504).

[0054] Step 508 is performed when unsettled stop Ward Laon is in the sentence extracted at the decision step 506. a run is short (Short) -- ** -- it regards -- having (step 510) -- as for the Short Laon counter, the increment only of 1 is carried out (step 512). A flow returns to the decision step of 506. If all stop Ward Laon is processed, as for a flow, return and the following sentence will be extracted by step 502.

[0055] In the decision step 510, when a run is long (long), the contents of the Short Laon counter are checked. When the Short Laon counter is 0, step 520 is arranged in a start register at Ward who continues just behind current stop Ward Laon. The Short Laon counter is set as 0 (step 522), and a flow returns to the decision step 506. When the Short Laon counter is not 0 (step 514), - register is arranged at Ward just in front of current stop Ward Laon (step 516). The contents of the Short Laon counter are memorized (step 518). Next, a start register is arranged at the word which continues just after a current run (step 520). After the Short Laon counter is set as 0, a flow returns to the decision step 506.

[0056] The following is two from the speech of Shevardnadze. The underline of one main track is pulled and, as for short stop Ward Laon, the underline of two main tracks is lengthened by long stop Ward Laon.

[0057]

[External Character 8]

Ex. 1 -   of perestroika, the ideas of renewal, the ideas of democracy, of democratization. We did great work on

Ex. 2 -   by the events of the first day of the start of the work of our Congress; by the pressing of a button the fate not only of a president but of perestroika and democratization.

[0058] In the first example, supposing "We" suits VANISSHU Liszt, the number of short stop Ward Laon strings will increase more. In the 2nd example, supposing stop Liszt does not have "first", the number of short stop Ward Laon strings will increase more.

[0059] Although this approach is considered to be leading at least with it being the same as that of two previous examples, once Long Short Laon is identified, it needs cautions further about of which sentence the set should be elected. It is quite effective to elect the whole sentence in which being contained in some sentences also includes a certain unit or two or more Long stop Ward Laon at least.

[0060] In the 1st above-mentioned example, the score of six in seven short stop Ward Laon was carried

h   g   cg b   eb cg e e

out to "1" in either the stop / the VANISSHU method or the Short method. Therefore, probably, election of the sentence corresponding to the approach was appropriate in all three examples.

[0061] Since it is used for abstract creation, it is possible to use the two or more above-mentioned procedures of electing a passionate sentence. Two approaches combine or even three combination of all approaches is easy. It is expectable for the statistical difference between the detailed properties for every approach that nearby becomes better than the case where only only one approach is used as for the approach put together.

[0062] The score of the sentence obtained using two approaches is combinable by adding, before selecting a passionate sentence. ranking of the sentence is carried out (or the value from which the rank was obtained -- generating), and another technique used before electing a passionate sentence adds the result together. Or in consideration of the Short approach, a stop and VANISSHU Liszt can change so that only a short word may be included, and either a stop / the VANISSHU approach or the Long Short approach is used further. Of course, it can also be used in order to carry out the score of the sentence for other combination of these approaches.

[0063] Furthermore, this invention is usable as a part of the automatic document proofreading (revision) tool which selects the sentence which was not elected as an abstract as a candidate who may be proofread. The sentence (not emphasized) which is not strong as for **** is elected for scrutinization, and it urges to a user, and proofreading of a sentence (minding a display unit 16) is made to be considered in emphasis detection. Or each sentence in a text can also attach a comment for emphasis (it has rough criteria shown [emphasis / a boldus face and / like inside] in the emphasis strongest against arbitration by change of font which makes fewest emphasis italic type in the usual state) rating (grading), and all the information on the others considered to be suitable (for example, length of a sentence (die length)).

[0064]

[Effect of the Invention] Through detection of the word and/or phrase which show emphasis, an opium poppy with a rank and an abstract can be created for the sentence of a document, or this invention can edit a document.

[Translation done.]

h   g   cg b   eb cg   e e

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] The step which is an approach for discriminating the field of a text from the electronic filing document using a digital computer automatically, and identifies the stop WORD in a document, The step which determines whether it is VANISSHU WORD any of the identified stop WORD they are, The automatic discernment approach equipped with the step which answers the stop WORD and VANISSHU WORD in each field, and carries out scoring of the field of a document, and the step which identifies a number of fields with which the document was beforehand defined based on the score of a field.

[Translation done.]

h   g   cg b   eb cg e e   h   g